# Excursion and contour uncertainty regions for latent Gaussian models

David Bolin[1] and Finn Lindgren[2]

[1]*Department of Mathematics and Mathematical Statistics, Umeå University, Umeå, Sweden*
[2]*Mathematical Sciences, University of Bath, Bath, United Kingdom*

**Abstract:** In several areas of application ranging from brain imaging to astrophysics and geostatistics, an important statistical problem is to find regions where the studied process exceeds a certain level. Estimating such regions so that the probability for exceeding the level in the entire set is equal to some predefined value is a difficult problem connected to the problem of multiple significance testing. In this work, a method for solving this problem, as well as the related problem of finding credible regions for contour curves, for latent Gaussian models is proposed. The method is based on using a parametric family for the excursion sets in combination with a sequential importance sampling method for estimating joint probabilities. The accuracy of the method is investigated using simulated data and an environmental application is presented.

**Key words:** Latent Gaussian models; excursion sets; contour curves; multiple testing

## 1 Introduction

In many applications of spatial statistics, one is interested in finding areas where the studied process exceeds a certain level or is significantly different from some reference level. A typical example is in studies of air pollution, where the interest is in testing if, and where, the pollution level exceeds some given limit value set by a regulatory agency (Cameletti et al., 2012), and similar examples can be found in a wide range of scientific fields including brain imaging (Marchini and Presanis, 2003) and astrophysics (Beaky et al., 1992). In spatio-temporal applications one might be interested in finding regions that have experienced significant changes over the studied time period. This is a common problem in climate science and the studied quantity can for example be temperature (Furrer et al., 2007), precipitation (Sain et al., 2011), or vegetation (Eklundh and Olsson, 2003; Bolin et al., 2009). A closely related problem is that of quantifying uncertainty of contour curves, defined as sets of locations where the process transitions between being below and being above specified levels. Despite the ubiquitous use of contour maps in graphical representations of spatial inference, this problem has received surprisingly little attention (Lindgren and Rychlik, 1995; Polfeldt, 1999), and no commonly used methods exist.

The quintessential problem for level exceedances, or excursions, is that one has observations $\mathbf{y}$ of some latent stochastic field $x(\mathbf{s})$ and wants to find a region $D$ such that, with a certain given probability, $x(\mathbf{s}) > u$ for all $\mathbf{s} \in D$ for a given level $u$. The traditional approach is to formulate a hypothesis testing problem and correct the family-wise error rate. The easiest, and most common, way of specifying $D$ is as the set where the marginal probabilities exceed a threshold,

$$D_m = \{\mathbf{s} : \mathsf{P}(x(\mathbf{s}) > u) \geq 1 - \alpha\}, \tag{1}$$

where the probability is taken under the conditional distribution for $x|\mathbf{y}$. The set can be calculated using pointwise hypothesis testing where $\alpha$ acts as the pointwise type 1 error parameter. The problem with this definition of $D$ is that of multiple hypothesis testing; the parameter $\alpha$ does not give us information about the family-wise error rate, and hence does not quantify the certainty of the level being exceeded at all points in the set simultaneously. That is, the probability $\mathsf{P}(x(\mathbf{s}) > u, \mathbf{s} \in D_m)$ is in general smaller than $1 - \alpha$. If order to control the joint probability of exceedance on $D$, the set construction procedure needs to be modified.

The more general problem of multiple hypothesis testing is an active research area and a number of solution approaches for various contexts exist. Most of these solutions are based on first calculating

the marginal probabilities $\mathsf{P}(x(\mathbf{s}) > u)$, then calculating a single threshold value $t$, and finally defining the exceedance region as $D = \{\mathbf{s} : \mathsf{P}(x(\mathbf{s}) > u) > t\}$. The methods differ in how the threshold $t$ is calculated, and can basically be divided into three main categories; type 1 error control thresholding, false discovery rate thresholding, and posterior probability thresholding (Marchini and Presanis, 2003). The most popular method is likely the method by Adler (1981) using the Euler characteristic of the latent field to control the family-wise error rate when defining the threshold $t$. Though this method is simple to use, one has to be careful to check whether the required assumptions are satisfied. Typically the method is accurate for large values of $u$, and the latent field is assumed to be stationary.

In spatial geostatistical settings, the relevant level $u$ may not be large, so methods based on asymptotic arguments when $u$ goes to infinity are unlikely to perform well. For example, in the problem of finding regions with significant changes in vegetation in the African Sahel studied by Eklundh and Olsson (2003) and Bolin et al. (2009), the threshold is zero. The method due to French and Sain (2013) handles arbitrary levels $u$ by finding a threshold using simulation based approach.

Although methods based on hypothesis testing are applicable in many frequentistic settings, they do not transfer to a Bayesian hierarchical and latent model framework. It is therefore desirable to formulate questions regarding excursion and contour curve uncertainty as properties of posterior distributions for random fields. For excursion sets, this is easily done using the joint probability for exceeding the level in the entire set. For contour curve analysis, Lindgren and Rychlik (1995) defined a contour uncertainty region as a union of cross-section intervals where each interval contains a single level crossing with probability $1 - \alpha$, without correcting for multiple testing. Here, the problem is instead reformulated as a dual problem to that of excursion sets, by noting that a credible region for a contour can be obtained as the complement set of the excursions above and below the contour level.

We focus on the problem where the latent spatial field $x(\mathbf{s})$ is Gaussian and measured at a set of irregular locations. This means that the posterior distribution $\pi(x|\mathbf{y})$ is non-stationary and typically non-Gaussian unless the measurements are Gaussian and the model parameters are known a priori. The computational method is based on using a parametric family for the excursion sets in combination with a sequential importance sampling method for estimating joint probabilities. For a specific choice of the parametric family, the method is largely equivalent to the thresholding methods mentioned above, with the important difference that the correct joint distribution is used when selecting the threshold. The method is also extended using more general parametric families.

The aim is to provide a unified statement of the excursion and contour uncertainty problems for random fields, and develop methods for practical computations, and the structure of the article is as follows. In Section 2, the problem is formulated and definitions for excursion sets and credible regions for contour curves are given. In Section 3, a method for estimating these sets is proposed. Estimating the sets is the most difficult problem as one easily runs into computational difficulties arising from having to evaluate high-dimensional integrals. In Section 4, the methods are tested on a few simulated examples to test the method's accuracy. An application to air pollution data from the North-Italian region Piemonte is presented in Section 5, and a second application to vegetation index data from the African Sahel region is included in the supplementary online material. Finally, a few remarks and comments are given in Section 6.

## 2 Problem formulation

There are a number of different ways one could formulate excursion sets, and not all of them are useful from a practical point of view. Hence, in this section we will formalise the problem and discuss how the results should be interpreted. More precisely we look at two connected problems. The first one is to find areas where a stochastic process exceeds a given level with some probability and the second one is to quantify the uncertainty in contour curves of stochastic fields.

Let $\Omega$ be the domain of interest, e.g. for a statistical analysis. The choice of $\Omega$ has fundamental implications for the interpretation of the excursion sets, since they will depend on joint properties of

the stochastic process across the entire domain. In order to avoid unnecessary theoretical complications, we assume that $\Omega$ is an open set with a well-defined area $|\Omega| < \infty$. If the domain of interest is not open, we take the interior and ignore the behaviour on the boundary.

First some notation for excursion sets of a function and contour sets is needed.

**Definition 2.1** (Excursion sets for functions)**.** Given a function $f(\mathbf{s})$, $\mathbf{s} \in \Omega$, the positive excursion set for a level $u$ is defined by $A_u^+(f) = \{\mathbf{s} \in \Omega; f(\mathbf{s}) > u\}$. The negative excursion set is similarly defined by $A_u^-(f) = \{\mathbf{s} \in \Omega; f(\mathbf{s}) < u\}$.

In a similar fashion one could now define the set of contour points for the level $u$ as the set of points $\mathbf{s}$ for which $f(\mathbf{s}) = u$; however, a contour curve consists not only of these points but also discontinuous crossings of the level $u$. In order to incorporate both continuous and discontinuous crossings, a contour point is defined as a point $\mathbf{s}$ such that in every neighbourhood $B$ of $\mathbf{s}$

$$\exists\, \mathbf{s}_1, \mathbf{s}_2 \in B : \mathbf{s}_1 \neq \mathbf{s}_2, f(\mathbf{s}_1) \geq u, f(\mathbf{s}_2) \leq u.$$

The set of all such points is the complement of the union of the interiors of the excursion sets.

**Definition 2.2** (Contour sets for functions)**.** Given a function $f(\mathbf{s})$, $\mathbf{s} \in \Omega$, the contour set $A_u^c$ for a level $u$ is given by $A_u^c(f) = (A_u^+(f)^o \cup A_u^-(f)^o)^c$, where $A^o$ is the interior of the set $A$ and $A^c$ is the complement.

*Remark* 1. Taking the interiors of the sets $A_u^+(f)$ and $A_u^-(f)$ is important. Consider for example the following function on $\Omega = (0, 1)$

$$f(s) = \begin{cases} -1 & 0 < s < 0.5 \\ 1 & 0.5 \leq s < 1. \end{cases}$$

In this case $A_0^+(f) \cup A_0^-(f) = \Omega$, so without taking the interiors of the sets $A_0^c(f)$ would be empty, and thus fail to include the discontinuous crossing at 0.5 in the contour set. This may seem as only a theoretical nicety, but the problem with discontinuous functions occurs frequently in environmental applications when discontinuous covariates are used for the mean value function of the field. This makes it essential to not treat contour sets as regions where the function values are close to a level, but rather as regions where level *crossings* occur.

The statistical problem is now to find a region $D$ such that the function $x(\mathbf{s})$ exceeds the level $u$ with a certain probability $1 - \alpha$ for all $\mathbf{s} \in D$. There might be many such regions, so if one is interested in a single answer one might look for the largest of these.

**Definition 2.3** (Excursion set)**.** Let $x(\mathbf{s})$, $\mathbf{s} \in \Omega$ be a random field (or process). The positive level $u$ excursion set with probability $1 - \alpha$ is defined by

$$E_{u,\alpha}^+(x) = \arg\max_D \{|D| : \mathsf{P}(D \subseteq A_u^+(x)) \geq 1 - \alpha\}. \tag{2}$$

The corresponding negative level $u$ excursion set, $E_{u,\alpha}^-$, is defined by replacing $A_u^+(x)$ with $A_u^-(x)$ in (2).

*Remark* 2. The set $E_{u,\alpha}^+(x)$ can also be formulated as the largest set $D$ for which $\mathsf{P}(\inf_{\mathbf{s}\in D} x(\mathbf{s}) \leq u) \leq \alpha$, which can be useful when calculating the set in practice. Also note that for deterministic functions $f$ one has $E_{u,\alpha}^+(f) = A_u^+(f)$ and $E_{u,\alpha}^-(f) = A_u^-(f)$ for any $\alpha \in [0, 1]$.

It is important to realise how the excursion set $E_{u,\alpha}^+(x)$ should be interpreted: It is the largest set so that with probability $1 - \alpha$ the level $u$ is exceeded *at all locations* in the set. Therefore, it is a smaller set than $D_m$ defined in (1), which is the set of points where the marginal probability for exceeding the level is at least $1 - \alpha$. The related concept of a *credible region* for excursions is the set that with probability $1 - \alpha$ contains *all excursions*. This set, given by $E_{u,\alpha}^-(x)^c$, is larger than $D_m$.

Which set one is interested in depends on the application, but it can be a good idea to calculate both to get a better understanding of the uncertainties in the problem.

In certain applications, one might be interested in joint positive and negative excursions from some level, for example when doing simultaneous regressions and one is interested in finding regions where the slopes are significantly different from zero (see the online supplementary material for a possible scenario of this kind).

**Definition 2.4** (Contour avoiding set)**.** Let $x(\mathbf{s})$, $\mathbf{s} \in \Omega$ be a random field. The pair of joint contour $u$ excursion sets with probability $1-\alpha$ is given by

$$(M^+_{u,\alpha}(x), M^-_{u,\alpha}(x)) = \underset{(D^+,D^-)}{\arg\max}\{|D^- \cup D^+| : \mathsf{P}(D^- \subseteq A^-_u(x),\, D^+ \subseteq A^+_u(x)) \geq 1-\alpha\},$$

where the sets $(D^+, D^-)$ are open. The contour avoiding set is defined as the union of these two sets, $E_{u,\alpha}(x) = M^+_{u,\alpha}(x) \cup M^-_{u,\alpha}(x)$.

The contour avoiding set can now be used to define a credible region for the contour set.

**Definition 2.5** (Credible regions for contour sets)**.** Let $x(\mathbf{s})$, $\mathbf{s} \in \Omega$ be a random field, and let $E_{u,\alpha}(x)$ be the contour avoiding set. The credible region for the level $u$ contour set is $E^c_{u,\alpha}(x) = (E_{u,\alpha}(x))^c$.

The interpretation of the credible region is important. The set $E^c_{u,\alpha}$ is the smallest set such that with probability $1-\alpha$ *all* level $u$ crossings of $x$ are in the set. One should note that this definition of the credible region for contour sets is different from some other definitions in the literature, e.g. that of Lindgren and Rychlik (1995).

It is somewhat unsatisfactory that the sets defined here are made unique by finding the largest set satisfying a certain restriction. The set $E^+_{u,\alpha}(x)$ is for example defined as the largest set $D$ satisfying $\mathsf{P}(D \subseteq A^+_u(x)) \geq 1-\alpha$, but there are also many other smaller sets satisfying the requirement, and these are not seen if only $E^+_{u,\alpha}(x)$ is reported. Also, if one wants to know where the field likely exceeds the level $u$, the set $E^+_{u,\alpha}(x)$ might not be sufficient since it does not provide any information about the locations not contained in the set. Therefore, it would be good to have something similar to $p$-values, i.e. the marginal probabilities of exceeding the level, but which can be interpreted simultaneously. To that end we introduce four functions as tools for computing and visualising the $\alpha$-indexed families of sets $E^\bullet_{u,\alpha}(x)$.

**Definition 2.6** (Excursion functions)**.** For a level $u$, the positive and negative excursion functions, contour avoidance function, and the contour function are defined as

$$F^+_u(\mathbf{s}) = \sup\{1-\alpha; \mathbf{s} \in E^+_{u,\alpha}\}, \qquad F^-_u(\mathbf{s}) = \sup\{1-\alpha; \mathbf{s} \in E^-_{u,\alpha}\},$$
$$F_u(\mathbf{s}) = \sup\{1-\alpha; \mathbf{s} \in E_{u,\alpha}\}, \qquad F^c_u(\mathbf{s}) = \sup\{\alpha; \mathbf{s} \in E^c_{u,\alpha}\}.$$

Note that $F^c_u(\mathbf{s}) = 1 - F_u(\mathbf{s})$ and that the functions take values between zero and one. Each set $E^\bullet_{u,\alpha}$ can be retrieved as the $1-\alpha$ or $\alpha$ excursion set of the function $F^\bullet_u(s)$.

# 3 Computations

So far, no assumptions have been made regarding the distribution of $x(\mathbf{s})$, but to be able to calculate the excursion sets in practice we will now restrict ourselves to the class of latent Gaussian models, which is a popular model class with many practical applications (see e.g. Rue et al., 2009). Thus, the following problem setup is assumed. Let $x(\mathbf{s})$ be a random field that can be written on the form

$$x(\mathbf{s}) = \sum_{i=1}^{k} \beta_i f_i(\mathbf{s}) + z(\mathbf{s})$$

where $f_i(\mathbf{s})$ are fixed effects and $z(\mathbf{s})$ is a zero mean random field with covariance parameters $\boldsymbol{\theta}_1$. Further assume that both $z(\mathbf{s})$ and the parameter vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)^\top$ are a priori Gaussian.

Let $\mathbf{y} = (y_1, \ldots, y_m)^\top$ be a vector of measurements of the latent field with some distribution $\pi(\mathbf{y}|\mathbf{x}_{obs}, \boldsymbol{\theta}_2)$, where $\mathbf{x}_{obs}$ is a vector containing the latent field evaluated at the measurement locations and $\boldsymbol{\theta}_2$ is a vector of parameters for the measurement distribution. Finally let $\mathbf{s}_1, \ldots, \mathbf{s}_n$ be the set of locations where predictions of the latent field should be calculated and let $\mathbf{x} = (x(\mathbf{s}_1), \ldots, x(\mathbf{s}_n))^\top$. The posterior distribution for $\mathbf{x}$ can then be written as

$$\pi(\mathbf{x}|\mathbf{y}) = \int \pi(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{y})\,\mathrm{d}\boldsymbol{\theta}, \tag{3}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$, and this is the distribution that should be used in the probability calculations when estimating the excursion sets.

There are now, in principle, two main problems that have to be solved in order to find the excursion sets, contour avoiding sets, or contour credible regions:

**integration** For excursion sets, calculate $\mathsf{P}(D \subseteq A_u^+(\mathbf{x}))$ or $\mathsf{P}(D \subseteq A_u^-(\mathbf{x}))$ for a given set $D$, or in the case of contour avoiding sets or credible regions for contour sets, calculate $\mathsf{P}(D^- \subseteq A_u^-(\mathbf{x}), D^+ \subseteq A_u^+(\mathbf{x}))$ for the pair of sets $(D^+, D^-)$.

**shape optimization** Find the largest region $D$ satisfying the required probability constraint.

Hence, given a method to solve each of the two problems, one could simply run the shape optimization algorithm and in each iteration calculate the required probability using the integration method. In theory there are no problems doing this, but in practice the integration method will be computationally demanding and it may not be feasible to use this strategy for applications where the dimension of $\mathbf{x}$ is large. Therefore, we instead propose a slightly different strategy that will minimize the number of calls to the integration method by solving the problem sequentially. We first outline the strategy in the simplest possible situation of Gaussian process with known parameters, which will be used as a basis for the latent Gaussian setting.

The method is based on using an increasing parametric family for the excursion or contour sets in combination with a sequential integration routine for calculating the probabilities. The advantage with using a sequential integration routine is that if the required probability has been calculated for some set $D_1$, then the calculation for a larger set $D_2 \supset D_1$ can be based on the result for $D_1$, resulting in large computational savings.

**Algorithm 3.1** (Calculating excursion sets using a one-parameter family)**.** *Assume that* $\mathbf{x}$ *is a Gaussian variable with known parameters, and that* $D(\rho)$ *is a parametric family for the possible excursion sets, such that* $D(\rho_1) \subseteq D(\rho_2)$ *if* $\rho_1 < \rho_2$*. The following strategy is then used to calculate* $E_{u,\alpha}^+(\mathbf{x})$*.*

1. *Find the reordering given by the order the nodes are added to* $D(\rho)$ *when* $\rho$ *is increased.*
2. *sequentially add nodes to the set* $D$ *according to the ordering given above and in each step update the probability* $\mathsf{P}(D \subseteq A_u^+(\mathbf{x}))$*. Stop as soon as this probability falls below* $1 - \alpha$*.*
3. $E_{u,\alpha}^+$ *is given by the last set* $D$ *for which* $\mathsf{P}(D \subseteq A_u^+(\mathbf{x})) \geq 1 - \alpha$*.*

The algorithm is written for the case of positive excursion sets, and to use it for the other cases one only has to change $\mathsf{P}(D \subseteq A_u^+(\mathbf{x}))$ to the required probability. Also note that the excursion function $F_u^+(\mathbf{s})$ is retrieved by setting $\alpha = 1$ and in each step saving the probabilities $\mathsf{P}(D \subseteq A_u^+(\mathbf{x}))$. Thus, the entire excursion function can be retrieved by running the algorithm once.

Algorithm 3.1 is based on using a one-parameter family for the excursion sets, and before going into detail on how to do the steps in the algorithm, we outline a simple generalization based on using a $k$-parameter family.

**Algorithm 3.2** (Calculating excursion sets using a $k$-parameter family)**.** *Assume that* $\mathbf{x}$ *is a Gaussian variable with known parameters, and that* $D(\boldsymbol{\nu}, \rho)$ *is a parametric family for the possible excursion sets, such that* $D(\boldsymbol{\nu}, \rho_1) \subseteq D(\boldsymbol{\nu}, \rho_2)$ *if* $\rho_1 < \rho_2$ *for a fixed* $\boldsymbol{\nu}$*. The following strategy is then used to calculate* $E^+_{u,\alpha}(\mathbf{x})$*.*

1. *Do optimization of the size of* $D(\boldsymbol{\nu}, \bullet)$ *over* $\boldsymbol{\nu}$*:*
   - *For the current value of* $\boldsymbol{\nu}$*, find the reordering given by the order the nodes are added to* $D(\rho)$ *when* $\rho$ *is increased.*
   - *sequentially add nodes to the set* $D$ *according to the ordering given above and in each step update the probability* $\mathsf{P}(D \subseteq A^+_u(\mathbf{x}))$*. Stop as soon as this probability falls below* $1 - \alpha$*.*
   - *return the last set* $D$ *for which* $\mathsf{P}(D \subseteq A^+_u(\mathbf{x})) \geq 1 - \alpha$*.*
2. $E^+_{u,\alpha}$ *is given by the largest set* $D$ *found in the optimization over* $\boldsymbol{\nu}$*.*

This algorithm is more computationally demanding as one has to do numerical optimization over $\boldsymbol{\nu}$, and one should therefore avoid using parametric families where $\boldsymbol{\nu}$ is high-dimensional. For two-parameter families, the optimization can be done using a Golden section search or a similar fast optimization procedure for one-dimensional problems.

Before considering the problem of calculating excursion sets under the full posterior distribution (3), we go into some details on how to do the Gaussian calculations in practice. The problem of sequential integration is discussed in Section 3.1. A few parametric families for excursion sets and level avoidance sets are introduced in Section 3.2, which also contains a discussion on how to optimally reorder the nodes while doing the calculations. The latent Gaussian setting is finally discussed in Section 3.3.

### 3.1 Gaussian probability calculations

For a Gaussian vector $\mathbf{x}$, the probabilities required for calculating the contour and excursion sets can all be written on the form

$$I(\mathbf{a}, \mathbf{b}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \int_{\mathbf{a} \leq \mathbf{x} \leq \mathbf{b}} \exp(-\frac{1}{2}\mathbf{x}^\top \boldsymbol{\Sigma}^{-1}\mathbf{x})\, \mathrm{d}\mathbf{x}, \tag{4}$$

where $\mathbf{a}$ and $\mathbf{b}$ are vectors depending on the mean value of $\mathbf{x}$, the domain $D$, and on $u$. For example, for a positive excursion set, one has $b_i = \infty$ and $a_i = u - \mu_i$ for all $i$ such that $\mathbf{s}_i \in D$ and $a_i = -\infty$ for all other $i$.

The simplest way of approximating (4) is to use Monte-Carlo (MC) integration. However, estimating the probability with any reasonable accuracy using standard MC integration is often too computationally expensive. Fortunately there exist a number of variance reduction techniques that can be used for increasing the efficiency of the MC integration.

A key step in many numerical integration techniques is to transform the integral to make it more suitable for numerical integration. Notably, Genz (1992) derived such a transformation for the Gaussian integral (4), though similar transformations have been suggested by other authors as well (see e.g. Geweke, 1991). Besides transforming the integral to the unit hyper cube, the transformation also achieves a separation of the variables so that the full problem can be calculated sequentially. The integral can then efficiently be evaluated using a quasi MC (QMC) method where the uniform random numbers in the ordinary MC integrator are replaced by some deterministic sequence of points chosen to reduce the probabilistic error bound of the crude MC integrator, see Genz and Bretz (2009) for details. A final variance reduction technique is to reorder the variables before calculating the integral, as first suggested by Schervish (1984) and later improved by Gibson et al. (1994). These reorderings can reduce the error by an order of magnitude, as shown by Genz and Bretz (2002). However, the reordering technique is not applicable in our situation since the reordering will be determined by a parametric family for the excursion sets.

### 3.1.1 Methods for Markov random fields

A common assumption in spatial statics and image analysis is that the latent field can be modelled, or approximated, using a Gaussian Markov random field (GMRF). See Rue and Held (2005) for an introduction to GMRFs, and note that GMRFs are used also for modelling in continuous space, for example using the SPDE approach by Lindgren et al. (2011). One of the motivating reasons for using GMRFs is that it reduces the computational cost for parameter estimation and spatial prediction, and because of this one would also like to utilize the Markov property in the probability calculations.

The main difference between latent GMRF models and standard Gaussian models is that the distribution is specified using the (sparse) precision matrix $\mathbf{Q}$ instead of the covariance matrix:

$$I(\mathbf{a}, \mathbf{b}, \mathbf{Q}) = \frac{|\mathbf{Q}|^{1/2}}{(2\pi)^{d/2}} \int_{\mathbf{a} \le \mathbf{x} \le \mathbf{b}} \exp(-\frac{1}{2}\mathbf{x}^\top \mathbf{Q}\mathbf{x})\, \mathrm{d}\mathbf{x}, \tag{5}$$

Using the transformation method by Genz (1992) on GMRF models is difficult without first inverting the precision matrix and then ignoring the sparsity of $\mathbf{Q}$ in the calculations. To take advantage of the sparsity of $\mathbf{Q}$ one can instead use the fact that any GMRF can be viewed as a non-homogeneous auto-regressive process defined backwards in the indices of $\mathbf{x}$ (see Rue and Held, 2005, Theorem 2.7), that is, if $\mathbf{x}$ is a GMRF with mean $\boldsymbol{\mu}$ and precision matrix $\mathbf{Q}$, then

$$x_i | x_{i+1}, \dots, x_n \sim \mathsf{N}\left(\mu_i - \frac{1}{L_{ii}} \sum_{j=i+1}^{n} L_{ji}(x_j - \mu_j), L_{ii}^{-2}\right), \tag{6}$$

where $L_{ij}$ are the elements of the Cholesky factor of $\mathbf{Q}$. Let $\tilde{\mu}_i = \frac{1}{L_{ii}} \sum_{j=i+1}^{n} L_{ji}(x_j - \mu_j)$ and let $\pi_0(x_i)$ denote density function for the Gaussian distribution $\mathsf{N}(0, L_{ii})$. The integral can then be written as

$$I(\mathbf{a}, \mathbf{b}, \mathbf{Q}) = \int_{a_1+\tilde{\mu}_1}^{b_1+\tilde{\mu}_1} \pi_0(x_1) \int_{a_2+\tilde{\mu}_2}^{b_2+\tilde{\mu}_2} \pi_0(x_2) \cdots \int_{a_{d-1}+\tilde{\mu}_{d-1}}^{b_{d-1}+\tilde{\mu}_{d-1}} \pi_0(x_{d-1}) \int_{a_d}^{b_d} \pi_0(x_d)\, \mathrm{d}\mathbf{x}$$

where, because of the Markov structure, $\tilde{\mu}_i$ only depends on the elements in $x_{\mathcal{N}_i \cap \{i+1:d\}}$, and $\mathcal{N}_i$ is the neighbourhood of $i$ in the graph of the GMRF.

If $\mathbf{Q}$ is a band-matrix, the integral can be efficiently calculated as a sequence of iterated one-dimensional integrals as discussed in Genz and Kahaner (1986). However, the band width of $\mathbf{L}$ will often be too large for this method to be efficient, and a better alternative is then of use a particle filter algorithm based on the GHK simulator (Geweke, 1991; Hajivassiliou, 1991; Keane, 1993). Denote the integral of the last $d - i$ components by $I_i$,

$$I_i = \int_{a_i+\tilde{\mu}_i}^{b_i+\tilde{\mu}_i} \pi_0(x_i) \cdots \int_{a_{d-1}+\tilde{\mu}_{d-1}}^{b_{d-1}+\tilde{\mu}_{d-1}} \pi_0(x_{d-1}) \int_{a_d}^{b_d} \pi_0(x_d)\, \mathrm{d}\mathbf{x}$$

and note that the integral is the normalizing constant to the truncated density

$$f_i(\mathbf{x}_{i:d}) \propto 1(\mathbf{a}_{i:d} < \mathbf{x}_{i:d} < \mathbf{b}_{i:d}) \pi(\mathbf{x}_{i:d}).$$

Start by calculating $I_d = \Phi(L_{ii}b_d) - \Phi(L_{ii}a_d)$, simulate $N$ samples $\{x_d^j\}_{j=1}^N$ from the truncated normal distribution $h_d(x_d) \propto 1(a_d < x_d < b_d)\pi_0(x_d)$, and set $w_d^j = I_d$. The integrals $I_{d-1}, \dots, I_1$ are now estimated sequentially using importance sampling. In the $i$th step, simulate $x_{d-1}^j$ from the truncated normal distribution $h_i(x_i | x_i^j) = 1(a_i + \tilde{\mu}_i < x_i < b_i + \tilde{\mu}_i)\pi_0(x_i)$ and set $\mathbf{x}_{i:d}^j = \{x_i^j, x_{i+1:d}^j\}$. The integral $I_i$ is estimated as $I_i \approx \frac{1}{N} \sum_{j=1}^N w_i^j$ where $w_i^j$ are the importance weights which are updated recursively through $w_i^j = \left[\Phi\left(L_{ii}(b_i + \tilde{\mu}_i)\right) - \Phi\left(L_{ii}(a_i + \tilde{\mu}_i)\right)\right] w_{i+1}^j$.

To reduce the variance of the estimator when the target probability is small, a resampling step can be performed after having calculated the weights $w_i^j$. This is a common strategy in particle filter techniques, and the sample $\{\mathbf{x}_{i:d}^j\}$ is then updated by selecting $N$ particles from the set, where $x_{i:d}^j$ is selected with probability $w_i^j / \sum_{k=1}^N w_i^k$. To avoid resampling too often, one can do the resampling only if some criterion is met, for example if the effective sample size is below some given threshold.

### 3.2 Parametric families

As discussed previously, we calculate the excursion sets by assuming a parametric form of the sets $D$. The parametric families will be based on the marginal quantiles of $x(\mathbf{s})$, $\mathsf{P}(x(\mathbf{s}) \leq q_\rho(\mathbf{s})) = \rho$, which are easy to calculate using only the marginal posterior distributions. The simplest one-parameter family based on the marginal quantiles is given in the following definition.

**Definition 3.3** (One-parameter family)**.** Let $q_\rho(\mathbf{s})$ be the marginal quantiles for $x(\mathbf{s})$, then a one-parameter family for the positive and negative $u$ excursion sets is given by

$$D_1^+(\rho) = \{\mathbf{s}; \mathsf{P}(x(\mathbf{s}) > u) \geq 1-\rho\} = \{\mathbf{s}; \mathsf{P}(x(\mathbf{s}) \leq u) \leq \rho\} = A_u^+(q_\rho),$$
$$D_1^-(\rho) = \{\mathbf{s}; \mathsf{P}(x(\mathbf{s}) < u) \geq 1-\rho\} = \{\mathbf{s}; \mathsf{P}(x(\mathbf{s}) \geq u) \leq \rho\} = A_u^-(q_{1-\rho}).$$

Using this one-parameter family in Algorithm 3.1 is equivalent to finding a threshold value for the marginal excursion probabilities to get the correct simultaneous significance level. It is thus similar to the thresholding algorithms discussed in Marchini and Presanis (2003) but with the important difference that the correct joint, often non-stationary, posterior distribution is used when finding the threshold.

Although the simple one-parameter family appears to perform well in practice, it is not guaranteed to contain the largest of all possible candidate sets. One approach to improving this is to extend the definition to a two-parameter family that also considers other levels in the excursion sets.

**Definition 3.4** (Two-parameter family)**.** Let $q_\rho(\mathbf{s})$ be the marginal quantiles for $x(\mathbf{s})$, then a two-parameter family for the positive and negative $u$ excursion sets is given by

$$D_1^+(v,\rho) = \{\mathbf{s}; \mathsf{P}(x(\mathbf{s}) > v) \geq 1-\rho\} = \{\mathbf{s}; \mathsf{P}(x(\mathbf{s}) \leq v) \leq \rho\} = A_v^+(q_\rho),$$
$$D_1^-(v,\rho) = \{\mathbf{s}; \mathsf{P}(x(\mathbf{s}) < v) \geq 1-\rho\} = \{\mathbf{s}; \mathsf{P}(x(\mathbf{s}) \geq v) \leq \rho\} = A_v^-(q_{1-\rho}).$$

The sets $D_1^+(v,\rho)$ and $D_1^-(v,\rho)$ are increasing in $\rho$ for a fixed $v$.

A drawback with these parametric families is that they do not take the spatial dependency of the data into account directly. An alternative is to include smoothed versions of the sets, $\{\mathbf{s}; p_i^\tau \geq 1-\rho\}$, where $p_i^\tau$ are parametrically smoothed marginal excursion probabilities $p_i = \mathsf{P}(x(\mathbf{s}_i) > u)$.

**Definition 3.5** (Two-parameter smoothing family)**.** Let $p_i^\tau$ be the smoothed marginal positive $u$ excursion probabilities, using a circular averaging filter with radius $\tau$. A two-parameter family for the positive and negative $u$ excursion sets is then given by

$$D_2^+(\tau,\rho) = \{\mathbf{s}; p_i^\tau \geq 1-\rho\},$$
$$D_2^-(\tau,\rho) = \{\mathbf{s}; p_i^\tau \leq \rho\}.$$

The parameter $\tau$ determines how close $p_i^\tau$ is to the original excursion probabilities. For $\tau = 0$, no smoothing is done and for a general $\tau$, $p_i^\tau$ is equal to the average of the marginal excursion probabilities in the disk with radius $\tau$ centred at $\mathbf{s}_i$. As $\tau$ increases $p_i^\tau$ becomes smoother and approaches a constant function equal to the average excursion probability. One could also use other types of parametric smoothers instead of the simple averaging filter.

Finally, a parametric family that can be used for estimating credible regions for contour curves can be obtained by using the one-parameter family for the pair of contour avoiding sets.

**Definition 3.6** (Parametric family for contour avoiding sets)**.** Let $D_1^+(\rho_1)$ and $D_1^-(\rho_2)$ be given by Definition 3.3. A two-parameter family for the pair of contour avoiding sets is obtained as $(D_1^+(\rho_1), D_1^-(\rho_2))$. A one-parameter family is obtained by requiring that $\rho_1 = \rho_2 = \rho$.

### 3.2.1 Domain bounds and reorderings

In the case of a GMRF posterior, it is desirable to make the Cholesky factor of the precision matrix as sparse as possible, because it reduces the number of floating point calculations that have to be done and reduces the error of the estimator. Reordering the nodes according to a parametric family does not guarantee good sparsity of the Cholesky factor, but the reordering can be improved by finding upper and lower bounds for the region. In this section, we give some simple upper and lower bounds for positive excursion sets, and the changes needed for the other situations are obvious.

The simplest upper bound for the region is $U_1 = \{\mathbf{s} : \mathsf{P}(x(\mathbf{s}) > u) \geq 1 - \alpha\}$, which is calculated using only the marginal probabilities. If $\mathbf{x}$ is a perfectly correlated variable one has $E^+_{u,\alpha} = U_1$, and for the general case one has $E^+_{u,\alpha} \subset U_1$ since all points not in $U_1$ have marginal probabilities lower than $1 - \alpha$ of exceeding the level $u$.

A simple lower bound is obtained using Boole's inequality as $L_1 = \{\mathbf{s} : \mathsf{P}(x(\mathbf{s}) > u) \geq 1 - \alpha/n\}$ where $n$ is the size of $\mathbf{x}$. In terms of multiple hypothesis testing, this lower bound is obtained from the classical Bonferroni correction method and an improved lower bound can be obtained using the Holm-Bonferroni method (Holm, 1979) as $L_2 = \{\mathbf{s} : p_{(k)} > 1 - \alpha/k\}$ where $p_{(k)}$ is the $k$th largest probability in the set $\{\mathsf{P}(x(\mathbf{s}_i) > u), i = 1, \ldots, n\}$. If the stochastic variables $x(\mathbf{s}_i)$ are independent, ones has $E^+_{u,\alpha} = L_2$, and in the general case one has $E^+_{u,\alpha} \supset L_2$.

The nodes can now be categorized into three classes, the first class contains the nodes included in the lower bound $L_2$, the second class contains the nodes in the set $U_1 \setminus L_2$ and the third class contains all other nodes. Since one knows that all nodes in $L_2$ will be included in $E^+_{u,\alpha}$, these can be reordered to maximize the sparsity of the Cholesky factor, for example using an approximate minimum degree permutation. The nodes in the second class are then added in the order determined by the parametric family. Finally, since the nodes in the third class will not be included in the domain, these can be reordered to maximize the sparsity or integrated out of the posterior distribution. Making the bounds more precise will improve the sparsity of the problem and therefore reduce the MC error and the computational complexity. In practice, this type of reordering can be obtained using the approximate minimum degree ordering algorithm with ordering constraints by Amestoy et al. (1996, 2004).

## 3.3 Handling the latent Gaussian setting

In this section, we extend the method for calculating the excursion sets in the Gaussian setting to the latent Gaussian setting with posterior (3). The INLA method by Rue et al. (2009) is a good alternative for estimating the posterior distributions $\pi(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})$ and $\pi(\boldsymbol{\theta}|\mathbf{y})$ in practice; however, the choice of method to use for estimating these distributions does not affect the excursion set estimation so any appropriate method can be used to estimate these distributions. Given estimates of the posterior distributions, we propose five different methods for calculating the excursion probabilities. The basic idea behind all these methods is to use Gaussian approximations of either the posterior $\pi(\mathbf{x}|\mathbf{y})$ or the conditional posterior $\pi(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})$. We denote such Gaussian approximations $\pi_G(\mathbf{x}|\mathbf{y})$ and $\pi_G(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})$ and note that $\pi(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}) = \pi_G(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})$ if the likelihood of the model is Gaussian. Thus the Gaussian approximation of the conditional posterior is only needed for non-Gaussian likelihoods, and in this situation, such approximations can be achieved by matching the modal configuration and the curvature at the mode of the distribution, as described in Rue et al. (2009). The resulting approximation is a Gaussian distribution with some mean $\mu^\star$ and precision matrix $\mathbf{Q} + \mathrm{diag}(\mathbf{c}^\star)$, see Rue et al. (2009) for details.

**EB (Empirical Bayes)** In the EB method, the excursion set is calculated under the conditional posterior $\pi_G(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}_0)$ where $\boldsymbol{\theta}_0$ for example is the maximum a posteriori estimate or the maximum likelihood estimate of $\boldsymbol{\theta}$. Thus, the parameter uncertainty is ignored and the methods for Gaussian distributions can be used directly.

**QC (Quantile Correction)** The QC method is based on using the EB method with a modification of the integration limits in the Gaussian integrals based on the marginal posteriors. For each $i$, $\mathsf{P}(x_i > a_i|\mathbf{y})$ is calculated and the lower limits $a_i$ in (5) are replaced with $\tilde{a}_i = \sigma_i \Phi^{-1}\left(1 - \mathsf{P}(x_i > a_i|\mathbf{y})\right)$, where $\sigma_i$ is the marginal standard deviation for $x_i|\mathbf{y}, \boldsymbol{\theta}_0$ and $\Phi$ denotes the standard Gaussian distribution function. Similarly, $\mathsf{P}(x_i < b_i|\mathbf{y})$ is calculated and the upper limits $b_i$ in are replaced with $\tilde{b}_i = \sigma_i \Phi^{-1}\left(\mathsf{P}(x_i < b_i|\mathbf{y})\right)$. One then has that $\mathsf{P}_G(x_i > \tilde{a}_i|\mathbf{y}, \boldsymbol{\theta}_0) = \mathsf{P}(x_i > a_i|\mathbf{y})$ and $\mathsf{P}_G(x_i < \tilde{b}_i|\mathbf{y}, \boldsymbol{\theta}_0) = \mathsf{P}(x_i < b_i|\mathbf{y})$, where $\mathsf{P}_G(\cdot|\mathbf{y}, \boldsymbol{\theta}_0)$ denotes the probability calculated under the distribution $\pi_G(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}_0)$. Thus, the QC method is exact if the components $x_i$ are independent.

**NI (Numerical Integration)** In the NI method, one numerically approximates the excursion function as $F_u^\bullet(\mathbf{s}) = \sum_{k=1}^K \lambda_k F_{u,k}^\bullet(\mathbf{s})$ where $F_{u,k}^\bullet(\mathbf{s})$ is the level $u$ excursion function calculated for the conditional posterior $\pi_G(\mathbf{x} \mid \mathbf{y}, \boldsymbol{\theta}_k)$ for a fixed parameter configuration $\boldsymbol{\theta}_k$. The configurations $\boldsymbol{\theta}_k$ in the hyper parameter space can, for example, be chosen as in the INLA method and the weights $\lambda_k$ are chosen proportional to $\pi(\boldsymbol{\theta}_k|\mathbf{y})$ Finally, the desired excursion set for a fixed $\alpha$ is retrieved as the excursion set $A_\alpha^+(F_u^\bullet)$ of the excursion function.

**NIQC (Numerical integration with Quantile Corrections)** The NIQC method is based on combining the QC method and the NI method. The NI method is used for calculating the excursion function, but for each parameter configuration, the QC method is used to modify the integration limits while calculating $F_{u,k}^\bullet(\mathbf{s})$. This method is thus only required for models with non-Gaussian likelihoods, where $\pi_G(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}_k) \neq \pi(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}_k)$. For each configuration $\boldsymbol{\theta}_k$, the modified limits $\tilde{a}_i$ and $\tilde{b}_i$ are calculated as in the QC method so that so that $\mathsf{P}_G(x_i > \tilde{a}_i|\mathbf{y}, \boldsymbol{\theta}_k) = \mathsf{P}(x_i > a_i|\mathbf{y})$ and $\mathsf{P}_G(x_i < \tilde{b}_i|\mathbf{y}, \boldsymbol{\theta}_k) = \mathsf{P}(x_i < b_i|\mathbf{y})$.

**iNIQC (improved NIQC)** The iNIQC method is equivalent to the NIQC method except that it uses different corrections for each configuration, based on the quantiles for the marginal posterior $\pi(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}_k)$. For configuration $k$, the modified limits $\tilde{a}_i$ and $\tilde{b}_i$ are calculated so that $\mathsf{P}_G(x_i > \tilde{a}_i|\mathbf{y}, \boldsymbol{\theta}_k) = \mathsf{P}(x_i > a_i|\mathbf{y}, \boldsymbol{\theta}_k)$ and $\mathsf{P}_G(x_i < \tilde{b}_i|\mathbf{y}, \boldsymbol{\theta}_k) = \mathsf{P}(x_i < b_i|\mathbf{y}, \boldsymbol{\theta}_k)$.

The EB method is the simplest, and may be sufficient in many situations. The QC method is based on correcting the limits of the integral so that the probability would be correct if the $x_i$'s were independent. This method is as easy to implement as the EB method and should perform better in most scenarios. The NI method is $K$ times more computationally demanding as the probability has to be calculated for each parameter configuration $\boldsymbol{\theta}_i$, but should also be the most exact method for problems with Gaussian likelihoods. If the number of parameters is small one can often obtain accurate results with only a few parameter configurations, but the accuracy of the estimator will depend on how these configurations are chosen.

For non-Gaussian likelihoods, the NIQC and iNIQC methods can be used for improved results. The iNIQC is slightly more computationally demanding since one has to calculate the marginal quantiles for each configuration but should also perform better in practice for models with non-Gaussian likelihoods. Another possibility for handling non-Gaussian likelihoods is to modify the sequential integration method. How to do this will depend on the posterior distribution and this is therefore something that would have to be done separately for each problem type. Genz and Bretz (2009) outline how the quasi MC methods can be extended to $t$-distributions, and the GHK-based particle filter method can be extended to other types of distributions as well. However, it is outside the scope of this article to go into further details on this option.

# 4 Tests on simulated data

In this section, two examples using simulated data are presented to illustrate the methods and test their accuracy. In the first example, we look at a problem in one dimension with known model parameters, where a latent Gaussian process with an exponential covariance is observed under Gaussian measurement noise. In the second example, we look at a spatial model where a latent Gaussian Matérn field is observed under Gaussian measurement noise. The model parameters are estimated from data and the three methods for handling the full posterior distribution are compared.

## 4.1 Example 1: 1d Gaussian data with known parameters

We begin with a simple one-dimensional example to illustrate the different sets we have previously defined. Let $x(s)$, $s \in [0, 2]$ be a Gaussian process with an exponential covariance function with scaling parameter $\lambda = 1$ and mean

$$\mu(s) = \begin{cases} s - 0.5 & \text{if } s < 1 \\ 1.5 - s & \text{if } s \geq 1. \end{cases}$$

We generate a trajectory from the model and observe it at 500 locations $s_1, \ldots, s_n$ drawn at random in the interval under additive Gaussian noise $\varepsilon_i \sim \mathsf{N}(0, \sigma^2)$. Given the observations, we do spatial prediction to 1000 equally spaced locations in the interval, and estimate the positive 0-excursion function $F_0^+(s)$ using the parametric family $D_1^+(0, \rho)$. In Figure 1, Panel (a), the marginal excursion probabilities $\mathsf{P}(x(s) > 0)$ are shown in light grey and $F_0^+(s)$ is shown in grey.

By definition, the positive level 0 excursion set, $E_{0,\alpha}^+(x)$, is obtained as $A_{1-\alpha}^+(F_0^+)$, and this set is shown for $\alpha = 0.05$ in grey in Figure 1 (b). The light grey set shows the upper bound $U_1$, obtained as an excursion set of the light grey function in Panel (a), the dark grey set shows the Holm-Bonferroni lower bound $L_2$, and the black curve is the expected value of $x(s)$ given the data. Note that $L_2 \subset E_{0,\alpha}^+(x) \subset U_1$.

We now want to verify that the estimated sets $E_{0,\alpha}^+(x)$ have correct excursion probabilities, i.e. that $\mathsf{P}(x(s) > 0, s \in E_{0,\alpha}^+(x)) = 1 - \alpha$. To that end, draw $N$ samples from $\pi(x|\mathbf{y}, \boldsymbol{\theta})$, and let $N_s$ denote the number of samples for which $\inf\{x(s), s \in E_{0,\alpha}^+(x)\} \geq 0$. A MC estimate of the excursion probability is then given by $\hat{p}(\alpha) = N_s/N$ and one should have $\hat{p}(\alpha) \approx 1 - \alpha$ if $E_{0,\alpha}^+(x)$ is correctly estimated. In Figure 2 (a), the difference $1 - \alpha - \hat{p}(\alpha)$ is shown as a function of $1 - \alpha$. The difference is calculated twice, using two different estimates $\hat{p}(\alpha)$, each based on $N = 50000$ samples. The difference is small for all values of $\alpha$ and is mostly caused by the MC error in the estimation of $\hat{p}(\alpha)$, which has nothing to do with the accuracy of the method. Thus the sets $E_{0,\alpha}^+$ indeed have the correct excursion probabilities.

Finally in Figure 2 (b), the 0-contour credible region $E_{0,0.05}^c(x)$ is shown. The set was estimated using the two-parameter family for level avoidance sets from Definition 3.6 and Algorithm 3.2. The complement of this set is the union of the contour avoiding sets $(M_{0,0.05}^-(x), M_{0,0.05}^+(x))$, which is the largest pair of sets $(D^+, D^-)$ satisfying $\mathsf{P}(D^- \subseteq A_u^-(x),\, D^+ \subseteq A_u^+(x)) \geq 0.95$.

## 4.2 Example 2: 2d Gaussian data with unknown parameters

In this example, we change to a spatial model and compare the methods described in Section 3.3 for handling the full posterior distribution (3) in the calculations.

Let $x(\mathbf{s})$, $\mathbf{s} \in [0, 10] \times [0, 10]$, be a Gaussian field with a constant mean $\mu = 0$ and a Matérn covariance function

$$C(\|\mathbf{h}\|) = \frac{2^{1-\nu}\phi^2}{(4\pi)^{\frac{d}{2}}\Gamma(\nu + \frac{d}{2})\kappa^{2\nu}}(\kappa\|\mathbf{h}\|)^\nu K_\nu(\kappa\|\mathbf{h}\|), \tag{7}$$

where $\nu$ is a shape parameter, $\kappa^2$ a scale parameter, $\phi^2$ a variance parameter, $K_\nu$ is a modified Bessel function of the second kind of order $\nu > 0$, and $\|\cdot\|$ denotes the Euclidean spatial distance. We use the

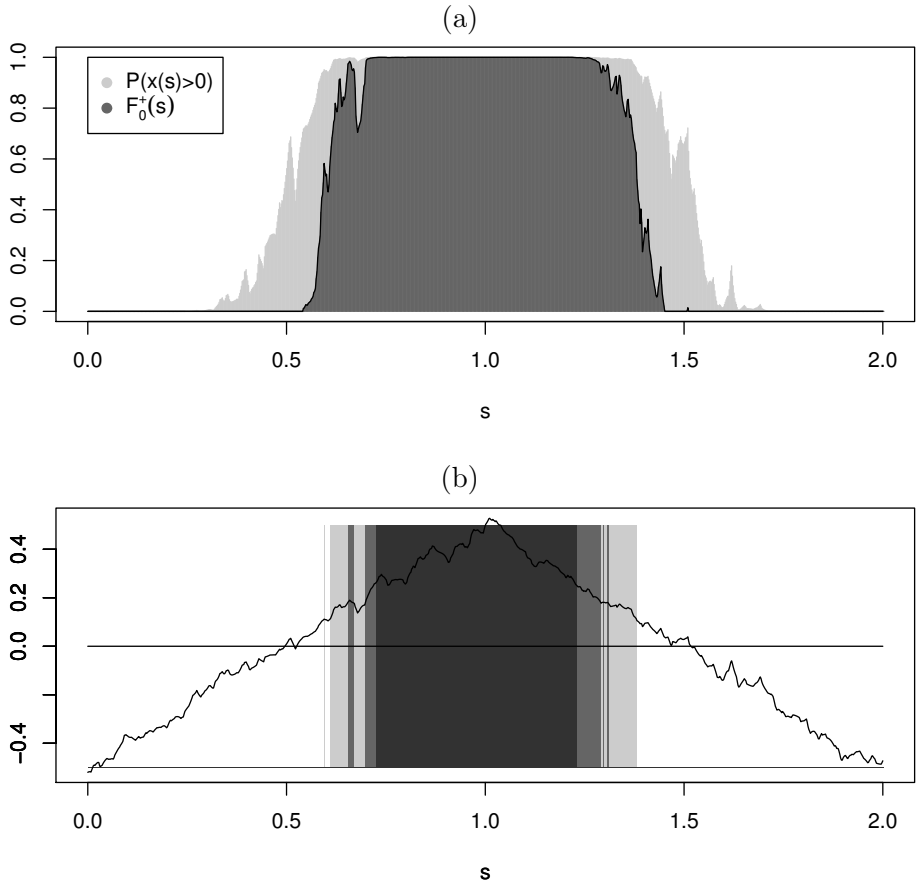


Figure 1: Results from Example 1. Panel (a) shows the marginal excursion probabilities $p(s) = \mathsf{P}(x(s) > 0)$ (light grey) and $F_0^+(s)$ (grey). Panel (b) shows $A_{0.95}^+(p)$ (light grey), which is the upper bound $U_1$, $E_{0,0.05}^+(x) = A_{0.95}^+(F_0^+)$ (grey), and the dark grey set is the lower bound $L_2$. The black curve is the kriging estimate of $x(s)$.

SPDE representation by Lindgren et al. (2011) of the field using a triangulation based on an $80 \times 80$ regular lattice in the region. The representation is a piecewise linear approximation $x(\mathbf{s}) \approx \sum_i x_i \varphi_i(\mathbf{s})$ of the field using 6400 piecewise linear functions $\varphi_i(\mathbf{s})$, each centred at one of the nodes in the lattice. The advantage with this representation is that it allows us to do all calculations using the weights $\mathbf{x}$ of the basis expansion, which form a GMRF.

We set $\nu = \phi = 1$, and $\kappa^2 = 2$, generate a realization of the field, and observe it under additive Gaussian noise $\mathsf{N}(0, 0.5^2)$ at 1000 locations chosen at random in the square. Given the measurements, the parameters and the marginal posterior distributions are estimated using the INLA method. The posterior estimate (kriging) of $\mathbf{x}|\mathbf{y}$ can be seen in the lower right panel of Figure 3, and the marginal probabilities $\mathsf{P}(x(\mathbf{s}) > 0|\mathbf{y})$ are shown in the lower left panel.

We now estimate the excursion function $F_0^+(\mathbf{s})$ using the three different methods described in Section 3.3 and the one-parameter family from Definition 3.3. The estimates can be seen in the upper panels of Figure 3. Visually it is in this case difficult to see any differences between the three estimates. To compare the accuracy of the estimates we do a comparison similar to the simulation study in Example 1, where MC simulation was used to estimate $\hat{p}(\alpha)$, the proportion of samples satisfying $\inf\{x(\mathbf{s}), \mathbf{s} \in E_{0,\alpha}^+(x)\} \geq 0$, which should be close to $1 - \alpha$ if $E_{0,\alpha}^+(x)$ is correct.

There are three possible sources of errors in this comparison. The first is the MC error from the estimation of $\hat{p}(\alpha)$, which has nothing to do with the accuracy of the method. The second is the MC error in the probability calculation when estimating the excursion distribution functions. This error is, however many orders of magnitude smaller in this case. The final source of error is the

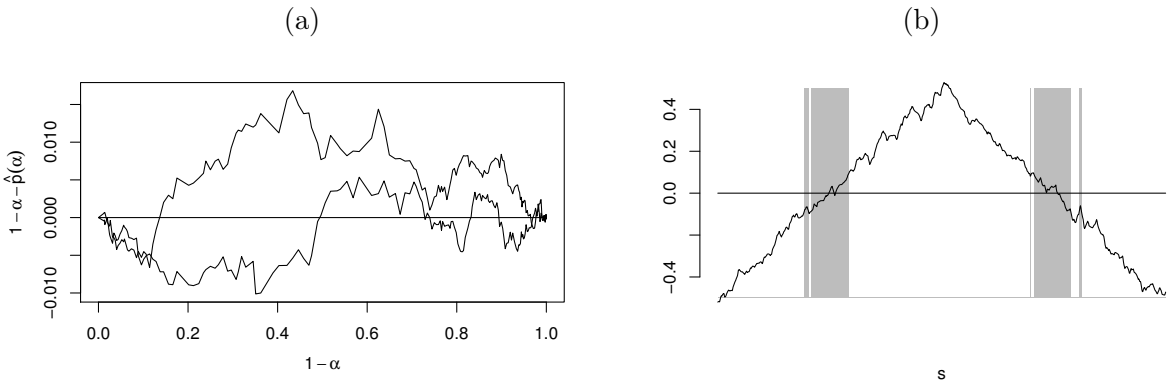


Figure 2: Results from Example 1. Panel (a) shows two estimates of $1-\alpha-\hat{p}(\alpha)$ as a function of $1-\alpha$. $\hat{p}(\alpha)$ is an estimate of $\mathsf{P}(x(\mathbf{s})>0, \mathbf{s}\in E^{+}_{0,\alpha}(x))$ based on MC simulation of $x(\mathbf{s})$, which should be close to $1-\alpha$ if $E^{+}_{0,\alpha}(x)$ is correctly estimated. The two curves show two results for two different MC simulations. Panel (b) shows the estimated contour credible region $E^{c}_{0,0.05}(x)$.

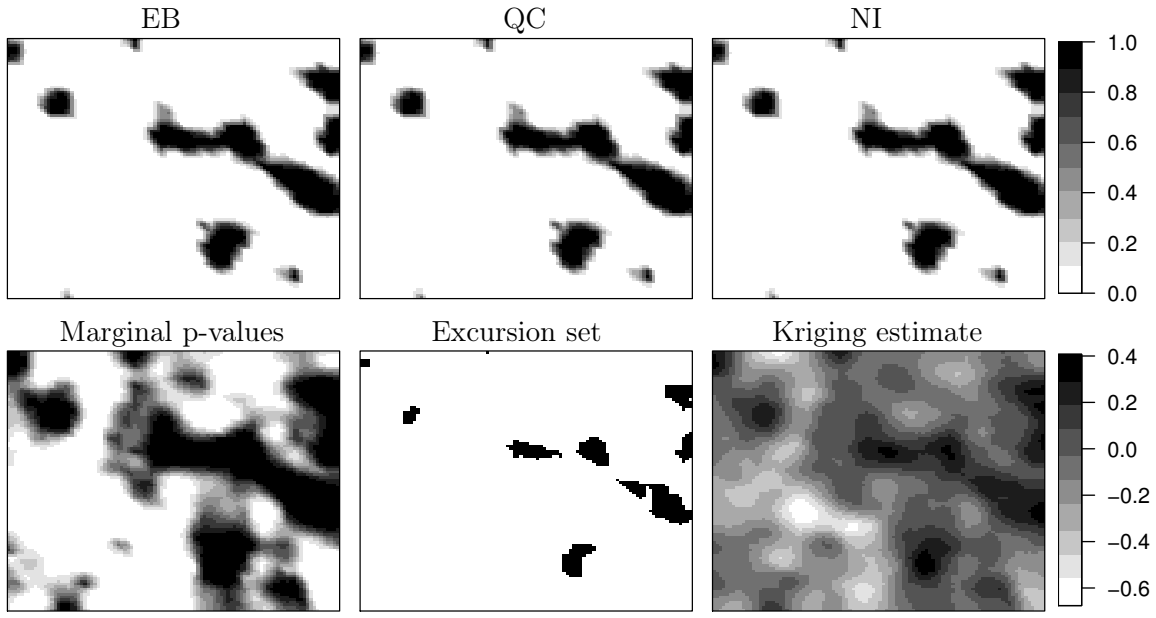


Figure 3: Results from Example 2. In the top row, three estimates of the excursion function can be seen using the EB method (left), the QC method (middle), and the NI method with 15 parameter configurations (right). In the bottom row, the marginal $p$-values for exceeding the limit can be seen in the left panel, using the same colour scale as for the top row. The middle panel shows the set $E^{+}_{0,0.05}(x)$ given by excursion function estimated by the NI method. Finally the right panel shows the kriging estimate of the latent field.

approximation error induced by using any of the three methods EB, QC, or NI for handling the full posterior distribution.

To investigate this approximation error, the difference $1-\alpha-\hat{p}(\alpha)$ is calculated for the three estimates of $F^{+}_{0}(\mathbf{s})$. The estimate of $\hat{p}(\alpha)$ is based on 20000 samples from the posterior $\pi(\mathbf{x}|\mathbf{y})$, obtained using an MCMC sampler described in the online supplementary material. In Figure 4 (a), the results can be seen for the EB method (dashed), the QC method (dash-dotted), the NI method with $k=45$ parameter settings (bright solid), and the NI method with $k=15$ parameter settings (dark solid). The comparison was done twice, with two different samples of size 20000 used for calculating $\hat{p}(\alpha)$, and the curves with the same line type show these two and give an indication of

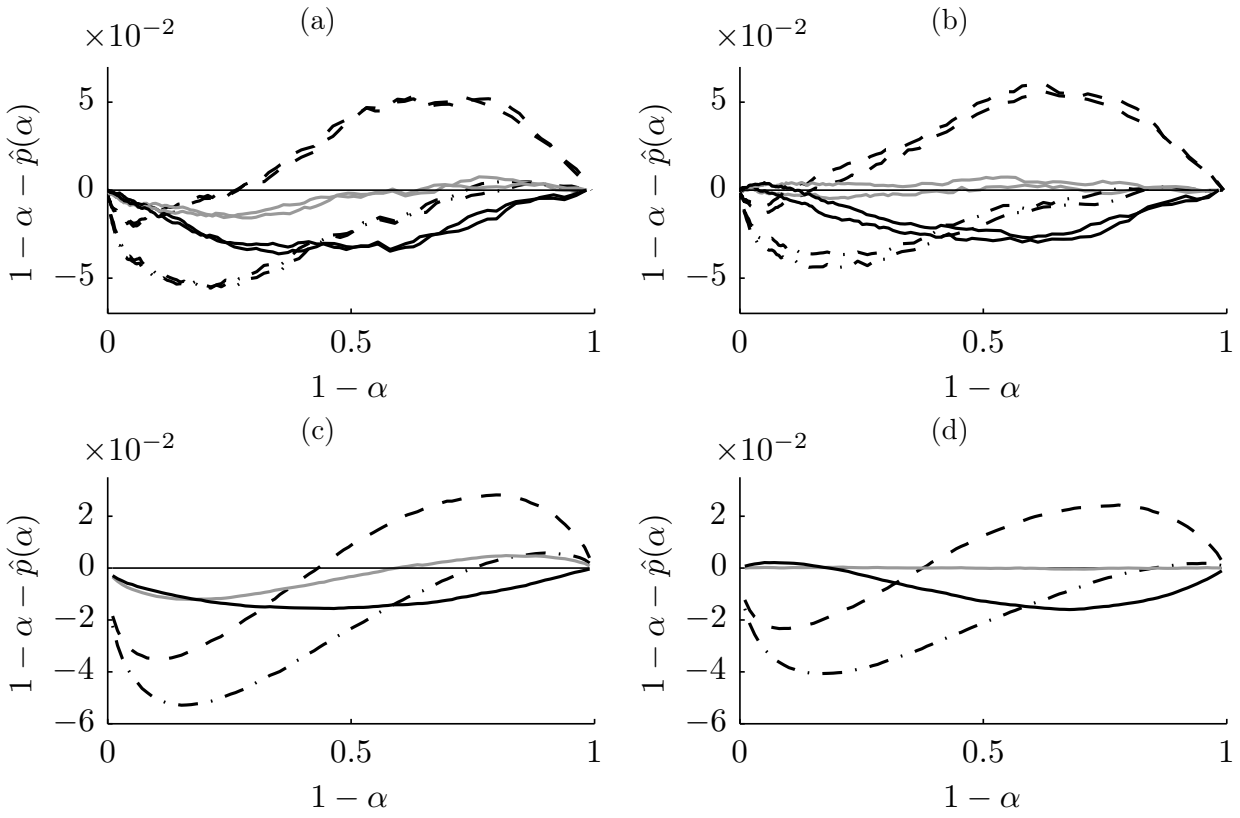


Figure 4: Results from Example 2 showing the difference $1-\alpha-\hat{p}(\alpha)$ as a function of $1-\alpha$ for the different approximation methods, EB (dashed), QC (dash-dotted), NI using 45 parameter configurations (bright solid), and NI using 15 parameter configurations (dark solid). $\hat{p}(\alpha)$ is an estimate of $\mathsf{P}(x(\mathbf{s})>0, \mathbf{s}\in E^+_{0,\alpha}(x))$ based on MCMC simulation of $x(\mathbf{s})$, which should be close to $1-\alpha$ if $E^+_{0,\alpha}(x)$ is correctly estimated. In the left panels, $\hat{p}(\alpha)$ is estimated using the full posterior distribution, and in the right panels, $\hat{p}(\alpha)$ is estimated using the discrete posterior distribution defined by the 45 parameter configurations from the NI method. In the upper panels, the comparison was done twice, with the two different estimates of $\hat{p}(\alpha)$, each based on 20000 samples of $x(\mathbf{s})$, and the curves of the same colour shows these two. In the lower panels, the average error from 50 different data sets is shown.

the size of the MC error in the comparison. As seen in the figure, the NI method performs best, as expected.

The error using the NI method comes from the fact that only finitely many points are used in the integration when approximating the posterior distribution for the parameters. That is, the full posterior $\pi(\boldsymbol{\theta}|\mathbf{y})$ is approximated by a discrete distribution with point masses at the parameter configurations $\boldsymbol{\theta}_i$. To verify that this indeed is the source of the error, we construct a second MCMC sampler where we instead of sampling $\boldsymbol{\theta}$ from the full posterior $\pi(\boldsymbol{\theta}|\mathbf{y})$ sample it from the discrete distribution defined by the 45 parameter configurations used in the first NI method. Panel (b) in Figure 4 shows the same comparison as Panel (a) but where $\boldsymbol{\theta}$ is sampled from the discrete distribution. As expected, the error for the NI method with 45 parameter configurations is now smaller.

The MC error from estimating $\hat{p}(\alpha)$ is quite large in Figure 4, so to get a better understanding of the other errors, a larger study was also performed where the procedure in Figure 4 was repeated 50 times for 50 different simulated data sets, and for each data set $N = 60000$ draws from the posterior was used when estimating $\hat{p}(\alpha)$. The average errors of these 50 runs can be seen in the lower panels of Figure 4. In Panel (c) the results using samples from the full posterior is shown, and in Panel (d) the results using the discrete distribution for $\boldsymbol{\theta}$ is shown. Note that the bright solid curve is very close to zero in Panel (c), indicating that the error in the NI method mostly depends on choosing the

integration points for $\boldsymbol{\theta}$ so that they capture the true posterior distribution well. Also note that the QC method performs well for large values of $1-\alpha$, and since one most often is interested in finding the excursion sets for small values of $\alpha$, this method is then a good way of finding such sets with less computational effort than using the NI method.

# 5 An application to air pollution data

High levels of air pollution can be harmful for the ecosystems and the human health. The effects on human health ranges from minor effects to the cardio-respiratory system to premature mortality (Cohen et al., 2009; Cameletti et al., 2012). Because of this, environmental agencies have to assess the air quality in order to take proper actions for improving the situation in polluted areas, and an important tool in this process is the ability to produce continuous maps of air pollution.

A region where the daily limit values fixed by the European Union for human health protection (see EU Council Directive 1999/30/EC) are periodically exceeded is the Piemonte region in northern Italy. Recently, Cameletti et al. (2012) proposed a statistical model to capture the complex spatio-temporal dynamics of $PM_{10}$ concentration in the region and used it to produce daily maps of $PM_{10}$. They also produced daily maps of exceedance probabilities of the value $50\mu g/m^3$, which is the value fixed by the European directive 2008/50/EC for the daily mean concentration that cannot be exceeded more than 35 days in a year. These probability maps only considered the marginal excursion probabilities, and no attempts of producing maps of simultaneous exceedance probabilities were made. In the following, we will therefore consider the same model and data but also estimate the excursion functions for the $50\mu g/m^3$ limit value.

Cameletti et al. (2012) considers daily $PM_{10}$ data measured at 24 monitoring stations by the Piemonte monitoring network during 182 days in the period October 2005 - March 2006, the data is provided by Aria Web Regione Piemonte. Denoting the measurements made at location $\mathbf{s}_i$ at time $t$ by $y(\mathbf{s}_i, t)$, the following measurement equation is assumed,

$$y(\mathbf{s}_i, t) = x(\mathbf{s}_i, t) + \varepsilon(\mathbf{s}_i, t), \tag{8}$$

where $\varepsilon(\mathbf{s}_i, t) \sim \mathsf{N}(0, \sigma_\varepsilon^2)$ is Gaussian measurement noise, both spatially and temporally uncorrelated, and $x(\mathbf{s}_i, t)$ is the latent field of true unobserved air pollution. The latent field is assumed to be on the form

$$x(\mathbf{s}_i, t) = \sum_{k=1}^{p} z_k(\mathbf{s}_i, t)\beta_k + \xi(\mathbf{s}_i, t), \tag{9}$$

where $p = 9$ covariates $z_k$ are used and $\xi$ is a spatio-temporal Gaussian random field. Based on the work of Cameletti et al. (2011) the following covariates were used: *1*) Daily mean wind speed; *2*) daily maximum mixing height; *3*) daily precipitation; *4*) daily mean temperature; *5*) daily emissions; *6*) altitude; *7*) longitude; *8*) latitude; and *9*) intercept. These covariates are provided with hourly temporal resolution on a 4 km $\times$ 4 km regular grid by the environmental agency of Piemonte region (Arpa Piemonte), see Finardi et al. (2008). The spatio-temporal process $\xi$ is assumed to follow first order autoregressive dynamics in time with spatially dependent innovations,

$$\xi(\mathbf{s}_i, t) = a\xi(\mathbf{s}_i, t-1) + \omega(\mathbf{s}_i, t), \tag{10}$$

where $|a| < 1$ and $\omega(\mathbf{s}_i, t)$ is a zero-mean temporally independent Gaussian process characterized by the spatio-temporal covariance function

$$\mathrm{Cov}(\omega(\mathbf{s}_i, t_1), \omega(\mathbf{s}_j, t_2)) = \begin{cases} 0 & \text{if } t_1 \neq t_2 \\ C(\|\mathbf{s}_i - \mathbf{s}_j\|) & \text{otherwise,} \end{cases} \tag{11}$$

where $C(\cdot)$ is a Matérn covariance function given by (7).

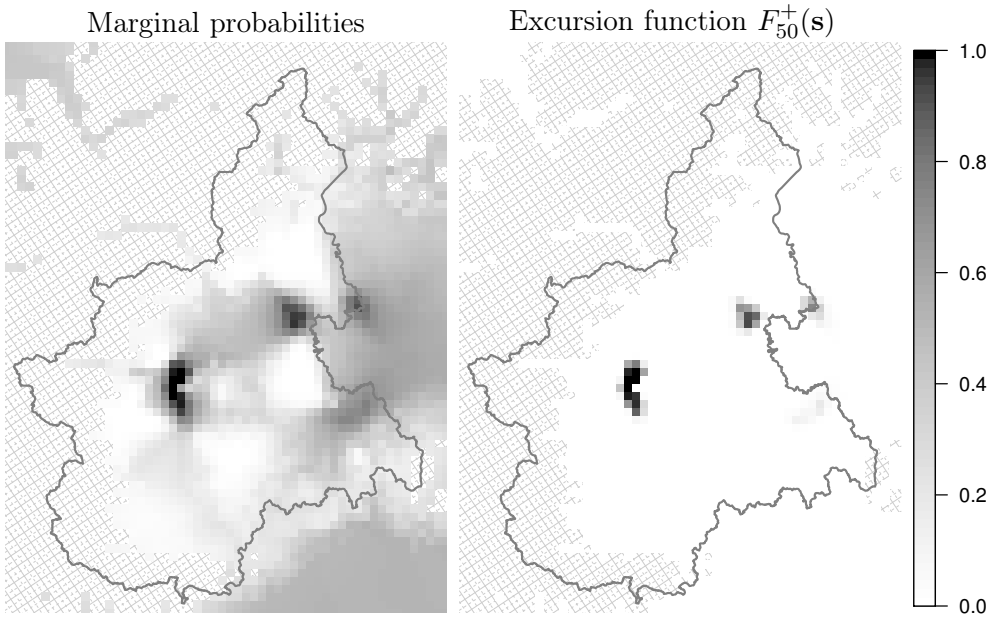


Figure 5: Results from the $PM_{10}$ application for January 30, 2006. A map of the marginal exceedance probabilities for $50\mu g/m^3$ (left), and the joint excursion distribution function for the level (right). The hashed region is excluded from the analyis.

The model parameters and the posterior distribution for the latent field are estimated using INLA in combination with the SPDE representation of Lindgren et al. (2011), see Cameletti et al. (2012) for details.

The map of marginal excursion probabilities for the level $50\mu g/m^3$ for January 30, 2006, based on the estimated posterior distribution for $x$, can be seen in the left panel of Figure 5. To avoid inappropriate linear extrapolation of the effect of elevation beyond the range of the elevation of the observations, the results are only based on areas below 1km. Based on these results, we now estimate the positive excursion function for the level $50\mu g/m^3$, $F_{50}^{+}(\mathbf{s})$, using the NI method from Section 3.3 and the parametric family of excursion sets from Definition 3.3. A total of 25 parameter configurations are used in the integration. The result can be seen in the right panel of Figure 5. As seen in the figure, there are three regions where the level is clearly exceeded, and a fourth that possibly contains too high pollution levels, although only two stand out clearly in the joint analysis. As expected, these areas coincide with the locations of the main metropolitan areas in the region; Turin, Novara, Vercelli, and Alessandria. One would like to make the predictions on a finer spatial scale, but since the covariates are given on a 4 km $\times$ 4 km grid, this spatial resolution has to be used.

To get a better understanding of the results, it is also of interest to find the regions where the pollution level is simultaneously below the limit value with some given probability. The marginal probabilities for being below be the level $50\mu g/m^3$, based on the estimated posterior distribution for $x$, can be seen in the left panel of Figure 6. The results are again only based on areas below 1km altitude. Using the same method as for the positive excursion function, we estimate the negative excursion function for the level $50\mu g/m^3$, $F_{50}^{-}(\mathbf{s})$, which can be seen in the right panel of Figure 6.

Note that the union of $E_{50,0.1}^{+}(x)$ and $E_{50,0.1}^{-}(x)$, shown in the left panel of Figure 7, covers only a small part of the region, indicating that the uncertainty in the problem is large. Also, the complement of the set $E_{50,0.1}^{-}(x)$, i.e. the union of the grey and black regions in the figure, contains all exceedances of the level with probability 0.9. This set is large, indicating that there are many regions where the level possibly is exceeded. Hence, it is important to note that the positive excursion set $E_{50,0.1}^{+}(x)$ is small because the uncertainty is large in the problem, and not because the other regions certainly have concentrations below the level.

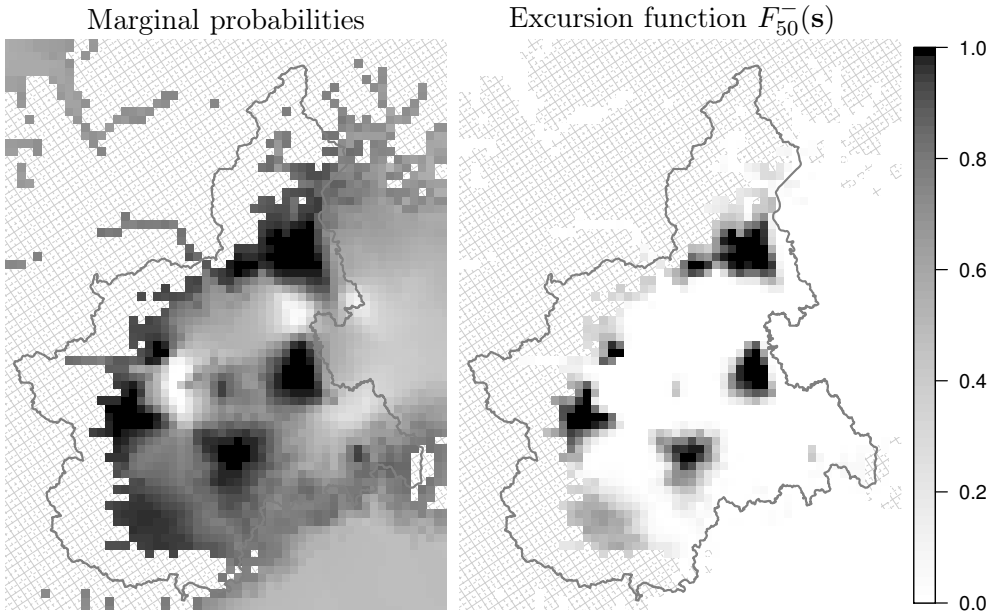


Figure 6: Results from the $PM_{10}$ application for January 30, 2006. A map of the marginal probabilities for the field being below the level $50\mu g/m^3$ (left), and the joint negative excursion distribution function for the level (right). The hashed region is excluded from the analyis.

To verify that the uncertainty is large, we finally calculate the contour function for the level $50\mu g/m^3$, $F_{50}^{c}(\mathbf{s})$, using the NI method and the one-parameter family from Definition 3.6 for the pair of level avoiding sets. The result can be seen in the right panel of Figure 7, and the resulting 90% credible region for the contour curve indeed covers a large part of the region which indicates that the uncertainty in the estimated contour curve is large.

# 6 Discussion

Estimating excursion sets and credible regions for contour curves for stochastic fields are difficult problems, both because of computational issues but also because it was not be clear how such uncertainty regions should be defined. In this work, we have given precise definitions for these regions, introduced the concept of excursion functions as a visual tool for illustrating the uncertainty in the regions, and presented a method for calculating these quantities for latent Gaussian models.

The main idea behind the computational method is to use a parametric family for the excursion sets in combination with a sequential integration method. Tests on simulated data showed that the method is accurate, and two applications (one in the online supplementary material) were presented to show that the method is applicable even to large environmental problems.

There are a number of extensions that could be made to this work. First of all, using the one-parameter family for the excursion sets gives a method that falls into the broad category of $p$-value thresholding methods for estimating simultaneous excursion sets. As previously mentioned, the important advantage with the method proposed here compared with other commonly used thresholding methods is that the correct joint distribution is used when selecting the threshold. The disadvantage is that the method is computationally more expensive than many of the standard thresholding methods. It would, therefore, be interesting to do a comparison with other similar methods with respect to the accuracy and computational complexity. Another interesting comparison would be to compare the credible regions for contour curves produced by these methods to those of Lindgren and Rychlik (1995). One could potentially also combine these methods with the work by Polfeldt (1999) to make

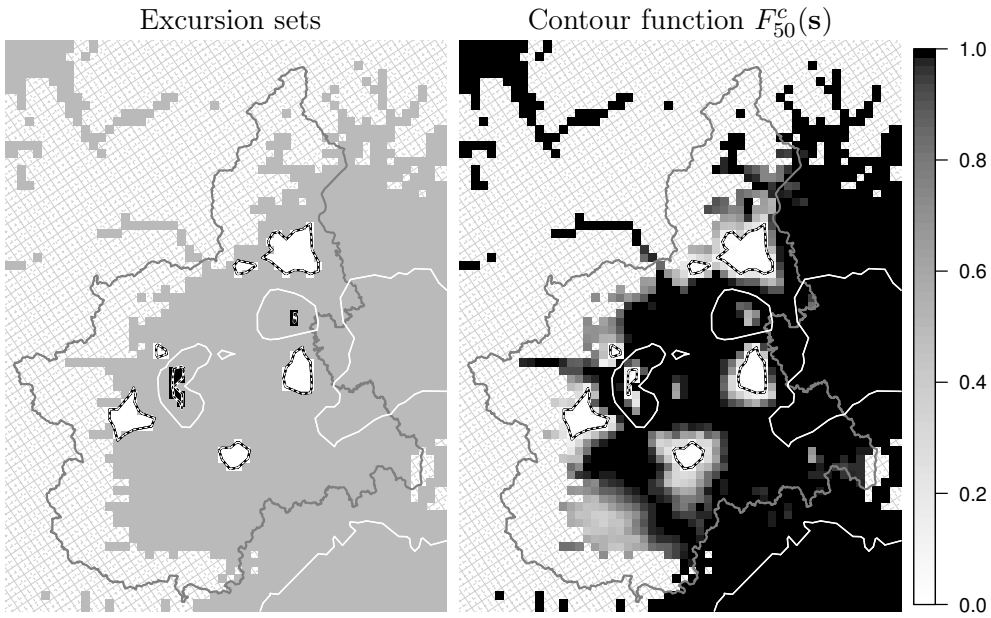


Figure 7: Results from the $PM_{10}$ application for January 30, 2006. In the left panel, the $E^+_{50,0.1}(x)$ and $E^-_{50,0.1}(x)$ excursion sets are shown in black and white, respectively, and their complement set in grey. The right panel shows the contour function for the level $50\mu g/m^3$ and the credible contour region $E^c_{50,0.1}(x)$ boundaries. In both panels, the contour is shown as a solid white curve. The hashed region is excluded from the analyis.

statements on the quality of contour maps.

We also presented a few two-parameter families that can be used for estimating the excursion sets, with the possibility of finding more precise estimates. Initial comparisons showed that there is not much gain in using these more complicated parametric families, but so far these comparisons have only been made using fairly simple latent models, and the gain is likely higher when the latent models are more complex. Hence, more studies are required to verify if this is the case and to investigate in what situations it is appropriate to use the simple one-parameter families. One possible advantage with the more complicated parametric families is when one has prior knowledge regarding the shape of the excursion sets. For example, if one knows that the excursion sets should be large contiguous regions, such knowledge could be incorporated using a two-parameter smoothing family.

More studies on how to best handle models with non-Gaussian likelihoods are also needed. A general framework for handling such models were presented, but the properties of the different methods should be investigated further.

As a final note, an R (R Core Team, 2013) package named `excursions` implementing the methods described in this work has been submitted to The Comprehensive R Archive Network (CRAN). This package also has an interface to R-INLA which can be used to analyze models estimated through the INLA method.

## Acknowledgements

The data used in the $PM_{10}$ study was provided by the information system Aria Web Regione Piemonte and Arpa Piemonte. The authors are grateful to Johan Lindström and Daniel Simpson for valuable discussions on the subject of excursions and contour curve uncertainty sets, and to Peter Guttorp for highlighting the need for a thorough treatment of the subject.

## References


Adler, R. J. (1981). *The Geometry of Random Fields*. Wiley, New York.

Amestoy, P., T. A. Davis, and I. S. Duff (1996). An approximate minimum degree ordering algorithm. *SIAM Journal on Matrix Analysis and Applications 17*(4), 886–905.

Amestoy, P., T. A. Davis, and I. S. Duff (2004). Algorithm 837: Amd, an approximate minimum degree ordering algorithm. *ACM Transactions on Mathematical Software 30*(3), 381–388.

Beaky, M. M., R. J. Scherrer, and J. V. Villumsen (1992). Topology of large-scale structure in seeded hot dark matter models. *Astrophys. J. 387*, 443–448.

Bolin, D., J. Lindström, L. Eklundh, and F. Lindgren (2009). Fast estimation of spatially dependent temporal vegetation trends using Gaussian Markov random fields. *Comput. Statist. and Data Anal. 53*, 2885–2896.

Cameletti, M., R. Ignaccolo, and S. Bande (2011). Comparing spatio-temporal models for particulate matter in piemonte. *Environmetrics 22*(8), 985–996.

Cameletti, M., F. Lindgren, D. Simpson, and H. Rue (2012). Spatio-temporal modeling of particulate matter concentration through the SPDE approach (submitted).

Cohen, M. A., S. D. Adar, R. W. Allen, E. Avol, C. L. Curl, T. Gould, D. Hardie, A. Ho, P. Kinney, T. V. Larson, P. Sampson, L. Sheppard, K. D. Stukovsky, S. S. Swan, L.-J. S. Liu, and J. D. Kaufman (2009). Approach to estimating participant pollutant exposures in the Multi-Ethnic Study of Atherosclerosis and air pollution (MESA air). *Environ. Sci. Technol. 43*(13), 4687–4693.

Eklundh, L. and L. Olsson (2003). Vegetation index trends for the African Sahel 1982–1999. *J. Geophys. Res. 30*, 1430–1433.

Finardi, S., R. De Maria, A. D'Allura, C. Cascone, G. Calori, and F. Lollobrigida (2008). A deterministic air quality forecasting system for Torino urban area, Italy. *Environmental Modelling and Software 23*(3), 344–355.

French, J. P. and S. R. Sain (2013). Spatio-temporal exceedance locations and confidence regions. *Annals of Applied Statistics. Prepress*.

Furrer, R., R. Knutti, S. R. Sain, D. Nychka, and M. G. A. (2007). Spatial patterns of probabilistic temperature change projections from a multivariate bayesian analysis. *Geophys. Res. Lett. 34*.

Genz, A. (1992). Numerical computation of multivariate normal probabilities. *J. Comput. Graph. Statist. 1*(2), pp. 141–149.

Genz, A. and F. Bretz (2002). Comparison of methods for the computation of multivariate t probabilities. *J. Comput. Graph. Statist. 11*(4), pp. 950–971.

Genz, A. and F. Bretz (2009). *Computation of Multivariate Normal and t Probabilities*, Volume 195 of *Lecture Notes in Statistics*. Springer.

Genz, A. and D. Kahaner (1986). The numerical evaluation of certain multivariate normal integrals. *J. Comput. Appl. Math. 16*, 255–258.

Geweke, J. (1991). Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities.

Gibson, G. J., C. Glasbey, and E. D.A. (1994). Monte Carlo evaluation of multivariate normal integrals and sensitivity to variate ordering. In I. Dimov, B. Sendov, and P. Vassilevski (Eds.), *Advances in Numerical Methods and Aplications*, pp. 120–126. World Scientific Publishing, River Edge.

Hajivassiliou, V. (1991, December). Simulation estimation methods for limited dependent variable models. Cowles Foundation Discussion Papers 1007, Cowles Foundation for Research in Economics, Yale University.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Statist. 6*(2), pp. 65–70.

Keane, M. (1993). 20 simulation estimation for panel data models with limited dependent variables. In *Econometrics*, Volume 11 of *Handbook of Statistics*, pp. 545 – 571. Elsevier.

Lindgren, F., H. Rue, and J. Lindström (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach (with discussion). *J. Roy. Statist. Soc. Ser. B Stat. Methodol. 73*, 423–498.

Lindgren, G. and I. Rychlik (1995). How reliable are contour curves? confidence sets for level contours. *Bernoulli 4*(1), 301–319.

Marchini, J. and A. Presanis (2003). Comparing methods of analyzing fMRI statistical parametric maps. *NeuroImage 22*, 1203–1213.

Polfeldt, T. (1999). On the quality of contour maps. *Environmetrics 10*, 785–790.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rue, H. and L. Held (2005). *Gaussian Markov Random Fields; Theory and Applications*, Volume 104 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC.

Rue, H., S. Martino, and N. Chopin (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *J. Roy. Statist. Soc. Ser. B Stat. Methodol. 71*(2), 319–392.

Sain, S. R., R. Furrer, and N. Cressie (2011). A spatial analysis of multivariate output from regional climate models. *Ann. Appl. Statist. 5*, 150–175.

Schervish, M. J. (1984). Algorithm as 195: Multivariate normal probabilities with error bound. *J. Roy. Statist. Soc. Ser. C Appl. Statist. 33*(1), pp. 81–94.